DESCRIPTION

METABOLIC SELECTION METHODS

FIELD OF THE INVENTION

The present invention relates to methods for screening for enzymatic pathways, and the isolation of the genes and proteins that make up these pathways.

BACKGROUND OF THE INVENTION

The following description of the background of the invention is provided to aid in understanding the invention, but is not admitted to be, or to describe, prior art to the invention.

Biological synthesis of compounds is frequently more cost effective and more productive than chemical synthesis, which can have low yields, require expensive and toxic reagents, and require lengthy purifications. In contrast, biological synthesis using known pathways can be rapid, with high yields. However, the identification of new biological pathways for syntheses of interest is difficult and time consuming.

Currently, the biochemical screening of isolates is a major means by which people find new pathways for the production of chemicals, antibacterials, and other anti-infectives. However, screening is inherently several orders of magnitude slower than selection and requires that the organism be cultured in the laboratory. Since at least 99%

of the microbes in the environment do not grow on laboratory media, less than 1% can be tested using a biochemical screen. Thus, biological pathways in 99% of organisms will never be found by classical biochemical screening technologies.

SUMMARY OF THE INVENTION

The metabolic selection strategy of this invention is designed to find an enzymatic pathway for the conversion of any source compound to any target compound. Conservatively, this technique allows at least a million-fold increase in the discovery rate over classical biochemical screening approaches, and allows testing of the 99% of the environmental microbes that are currently unable to be cultured in the laboratory.

A biocatalytic or metabolic pathway consists of a series of protein catalysts (enzymes) which catalyze the conversion of a starting material to the final product. A general process to identify the metabolic pathway from a source compound to a target compound involves the creation/identification of an easily genetically-manipulatable organism containing an inducible signal, which is activated when a target compound is metabolized. This is followed by the screening of nucleic acid in this organism to identify genes which metabolize the source compound to the target compound.

An example of a selection strategy which can be used to identify the metabolic pathway from a source compound to a target compound is diagrammed in Figure 11. As a first step, microbial isolates are selected that are capable of metabolizing a target compound "T", but not a source compound "S", to an essential factor. Essential factors can include elements like carbon, sulfur, phosphorous, and nitrogen, or other essential nutrients, *e.g.* some amino acids, fatty acids, and carbohydrates. In a second step, the pathway responsible for the catabolism of compound "T" is identified and made conditional. That is, the gene(s) for the pathway is cloned and placed under control of an inducible promoter such that growth on the target compound is turned "ON" only when the inducer is present. This engineered strain is referred to as the "tester strain". The third part of the strategy is the transfer of foreign DNA from environmental sources into the tester strain, followed by selection for growth on the source compound "S" in the presence of inducer. Such positive clones either are capable of metabolizing compound "S" in the absence of inducer, in which case utilization of "S" does not require prior conversion to compound "T" (Figure 11; pathway I), or alternatively metabolize compound "S" only when "T" catabolism is "ON", suggesting that utilization of "S" proceeds via compound "T" to intermediary metabolism (Figure 11; pathway II). These latter clones are further analyzed and the biocatalysts for the conversion of

"S" to "T" are characterized. A specific embodiment of the metabolic selection strategy is shown in Figure 12, where "S" is 2-keto-L-gulonate (2-KLG), and "T" is ascorbic acid (AsA) which can be metabolized to carbon and energy.

Thus, in a first aspect, the invention features a method of screening for one or more nucleic acid sequences which express a product or products that convert a source compound into a target compound. The method comprises contacting a cell with one or more test nucleic acid sequences, where the cell expresses one or more genes encoding one or more proteins which, in the presence of the target compound, provide a detectable signal. The detectable signal indicates the presence of the desired nucleic acid sequence or sequences.

The term "screening" as used herein refers to methods for identifying a nucleic acid sequence of interest. Preferably, the method permits the identification of a nucleic acid sequence of interest among one or more sequences, more preferably among hundreds (100, 200,...900), most preferably among thousands (1,000, 2,000,...etc.) or more. The sequences to be screened can be isolated from one or more organisms. Preferably, the sequences are isolated from hundreds of organisms, more preferably from thousands or more organisms. The term "screening" may include both classical screening, whereby expression of the nucleic acid results in a phenotype that can be identified (for example by having a colony with the nucleic acid of interest change

color, fluoresce, or luminesce), and may also include classical selection, where typically the phenotype to be identified is growth on selective media. By "selective" is meant media on which the host strain will not grow or grows poorly, but that strains with the nucleic acid of interest will grow in a manner which can be readily distinguished from host strain growth by methods well-known in the art.

The term "nucleic acid" as used herein refers to either deoxyribonucleic acid or ribonucleic acid that may be isolated, enriched, or purified from natural sources or synthesized recombinantly. These methods are well-known in the art and specific examples are also given herein. Preferably, a "nucleic acid" to be identified in the screening method comprises a nucleic acid encoding a metabolic pathway that is not normally found in the cell. Thus, preferably, the pathway has not simply been inactivated through a mutation and the relevant genes are now being identified through complementation. Rather the nucleic acid being identified does not normally exist in the cell in which it is being screened for. Typically, the screening is cross strains, more typically, cross-species, and even more preferably, cross-genera or with further remoteness.

By "isolated, purified, or enriched" in reference to nucleic acid is meant a polymer of 6 (preferably 21, more preferably 39, most preferably 75) or more nucleotides conjugated to each other, including DNA and RNA that is

isolated from a natural source or that is synthesized. In certain embodiments of the invention, longer nucleic acids are preferred, for example those of 300, 600, 900 or more nucleotides and/or those having at least 50%, 60%, 75%, 90%, 95% or 99% identity to the sequence shown in SEQ ID NO:1; SEQ ID NO:2; SEQ ID NO:3; SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, or SEQ ID NO:19.

The isolated nucleic acid of the present invention is unique in the sense that it is not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular (*i.e.*, chromosomal) environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide chain present, but that it is essentially free (about 90-95% pure at least) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

By the use of the term "enriched" in reference to nucleic acid is meant that the specific DNA or RNA sequence constitutes a significantly higher fraction (2-5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other

DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted that "enriched" does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term "significant" is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other nucleic acids of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no DNA or RNA from other sources. The other source DNA may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes from naturally occurring events, such as viral infection, or tumor type growths, in which the level of one mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment (compared to the natural level this level should be at least

2-5 fold greater, *e.g.*, in terms of mg/mL). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^6-fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

The term "expresses a product" as used herein refers to the production of proteins from a nucleic acid vector containing genes within a cell. The nucleic acid vector is transfected into cells using well known techniques in the art as described herein. The "product" may, or may not, be naturally present in the cell.

The term "nucleic acid vector" relates to a single- or double-stranded circular nucleic acid molecule that can be transfected into cells and replicated within or

independently of a cell genome. A circular double-stranded nucleic acid molecule can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a desired product can be inserted into a vector by cutting the vector with restriction enzymes and ligating the pieces together, depending on the availability of useful restriction sites. However, there are many methods well-known in the art for the insertion of nucleic acid sequences into vectors.

The term "transfecting" as used herein includes a number of methods to insert a nucleic acid vector or other nucleic acid molecules into a cellular organism. These methods involve a variety of techniques, such as treating the cells with high concentrations of salt, an electric field, detergent, or DMSO to render the outer membrane or wall of the cells permeable to nucleic acid molecules of interest or use of various viral transduction strategies.

The term "converts" as used herein refers to changing one compound into another compound, preferably enzymatically. The "source compound" refers to the compound to be converted to the "target compound." The "target compound" includes not only the compound that is metabolized to form a detectable signal, but can also include intermediates along the path to a detectable signal. This

is particularly preferred if the target compound is a surrogate target. By "surrogate target compound" is meant a target that is used because the preferable target cannot be used for any of several potential reasons (*e.g.* if it doesn't cross membranes, has a short half-life, easily broken down, etc.). The "target compound" also includes interconvertible compounds. By " interconvertible" is meant that a pathway exists in the tester strain to convert the compound to the target compound.

The term "contacting" as used herein refers to mixing a solution comprising the test nucleic acid with a liquid medium bathing the cells of the methods. The solution comprising the nucleic acid may also comprise other components, such as dimethyl sulfoxide (DMSO), which facilitates the uptake of the test nucleic acid into the cells of the methods. This may also be done by other methods well-known in the art including, but not limited to, transfection or transformation techniques. The solution comprising the test nucleic acid may be added to the medium bathing the cells by utilizing a delivery apparatus, such as a pipet-based device or syringe-based device.

The term "cell" as used herein includes the typical definition of a cell, and is further specifically intended to include "cell-free" systems comprising the cellular machinery necessary to express the nucleic acid of the invention. By "cellular machinery" is meant the cellular components present in cell-free transcription

and/or translation systems. Such systems are well-known in the art. In particular, the "cell" lacks the ability to convert a source compound into a target compound, prior to the addition of test nucleic acid sequences. The term "lacks the ability" also includes cells in which the activity may be present but is at too low a level to provide a detectable signal, or is low enough that an additional activity is detectably different. By "detectably different" is meant able to be measured over the background level (*e.g.* the level of the signal endogenously present in the "cell" and in the equipment used to measure the signal) by an amount greater than the level of error present in the method of measuring.

The term "detectable signal" as used herein refers to a method of identification of the nucleic acids of interest *e.g.* by color, fluorescence, luminescence or growth.

In preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the one or more nucleic acid sequences encodes a metabolic pathway not normally present in said cell. A "metabolic pathway" consists of a series of protein catalysts (enzymes) which catalyze the conversion of a starting material to a product. And further, by "metabolic pathway" is meant the enzymes, and genes that encode them, that metabolize a source compound to a target compound.

In other preferred embodiments, the nucleic acid is selected from the group consisting of mutagenized DNA, environmental DNA, combinatorial libraries, and recombinant DNA. Preferably, the environmental DNA is selected from the group consisting of mud, soil, sewage, flood control channels, sand, and water. Preferably the mutagenized DNA is the result of enzyme mutagenesis where the mutagenesis is selected from the group consisting of random, chemical, PCR-based, and directed mutagenesis. The directed mutagenesis is to include, for example, DNA shuffling. Preferably the enzymes to be mutagenized in this way are selected from the group consisting of lactonases, esterhydrolases, and reductases.

The term "environmental" as used herein refers to nucleic acids extracted from the environment, *e.g.* from mud, soil, or water. By "extracted" is meant isolated, enriched, or purified as defined above. The environmental sample can be directly extracted without prior laboratory culture, or can be pre-cultured, for example, in the presence of a growth selective agent. Methods are known in the art and examples are described herein.

In still other preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color. Methods for detecting these signals are well-known in the art. Preferably, the

detectable signal is growth, and the target compound provides an element or factor required for growth. Preferably the target compound is selected from the group consisting of ascorbate and 2-keto-L-gulonate (2-KLG), most preferably ascorbate. Preferably the element is selected from the group consisting of carbon, nitrogen, sulfur, and phosphorous. Most preferably, the element is carbon. Alternatively, the essential factor is another essential nutrient. By "required for growth" is meant that the organism does not grow detectably in the absence of the element. By "provides an element" is meant that the compound can be metabolized by the organism, and that the result of this metabolism is the element in some form, *e.g.* carbon or carbon dioxide.

In other preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the source compound is selected from the group consisting of 2-keto-L-gulonate (2-KLG), 2,5-deoxy-keto-gulonate (2,5-DKG), L-idonate (L-IA), L-gulonate (L-GuA), and glucose, and most preferably 2-KLG.

In still other preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the cell naturally expresses the one or more genes encoding one or more proteins that in the presence of the target compound provide a detectable signal. Alternatively, the cell can be genetically manipulated to express the one or more genes encoding one or more proteins that in the presence of the target compound provide a detectable signal. In both cases, the one or more proteins are preferably Yia operon-related polypeptides. The one or more genes are preferably under the control of an inducible promoter. The inducible promoter preferably comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the $lacI^q$ repressor.

By "naturally expresses" is meant that the genes encoding the proteins are present in the cell in its natural state, *e.g.* in nature, prior to culture in the laboratory. The genes may or may not be expressed in the natural state, or may or may not be expressed constitutively or inducibly. By "genetically manipulated to express" is meant the transfection of the desired genes into the cell by methods well-known in the art, examples of which are described herein.

The term "promoter" as used herein, refers to nucleic acid sequence needed for gene sequence expression. Promoter regions vary from organism to organism, but are well known to persons skilled in the art for different

organisms. For example, in prokaryotes, the promoter region contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, ribosome binding site, start codon, and the like. By "inducible promoter" is meant a promoter which is only "on" in the presence of an inducer. The "inducer" is typically a small molecule. Inducible promoters and inducers are well-known in the art and examples are given herein.

The term "Yia operon-related polypeptides" as used herein refers to polypeptides comprising 12 (preferably 15, more preferably 20, most preferably 30) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; 31 (preferably 35, more preferably 40, most preferably 50) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; 5 (preferably 10, more preferably 15, most preferably 25) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; 17 (preferably 20, more preferably 25, most preferably 35) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; 11 (preferably 15, more preferably

20, most preferably 30) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:16; or a functional derivative thereof as described herein. In certain aspects, polypeptides of 100, 200, 300 or more amino acids are preferred. The Yia operon-related polypeptide can be encoded by its corresponding full-length nucleic acid sequence or any portion of its corresponding full-length nucleic acid sequence, so long as a functional activity of the polypeptide is retained (*see*, Examples section). It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide which retains the functionality of the original. In both cases, all permutations are within the embodiments of the invention.

The amino acid sequence of the Yia operon-related polypeptide will be substantially similar to the sequence shown in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18, or fragments thereof. A sequence that is substantially similar to the sequence of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18 will preferably have at least 90% identity (more preferably at least 95% and most preferably 98-100%) to the

sequence of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18 using a Smith-Waterman protein-protein search.

By "identity" is meant a property of sequences that measures their similarity or relationship. Identity is measured by dividing the number of identical residues by the total number of residues and gaps and multiplying the product by 100. "Gaps" are spaces in an alignment that are the result of additions or deletions of amino acids. Thus, two copies of exactly the same sequence have 100% identity, but sequences that are less highly conserved, and have deletions, additions, or replacements, may have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity. For example, the computer algorithm BLAST is preferably used to search for homologous sequences in a database, and CLUSTAL is used to perform alignments. Identity and similarity determinations can be made using a Smith-Waterman protein-protein search, for example.

In still other preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the cell grows on ascorbate and does not grow on 2-KLG. Alternatively, the cell may grow on 2-KLG and not grow on 2,5-DKG. Preferably the cells are bacteria. Most preferably, the cell selective for ascorbate is *Klebsiella oxytoca*. By "grows on" is meant that the cell

can utilize the compound (*e.g.* ascorbate or 2-KLG) as a source of carbon in the minimal essential media. However, the cell is unable to grow in the minimal essential media in the absence of the provided carbon source. Thus, this provides a selective tool for the identification of the nucleic acid encoding the polypeptides of interest.

A second aspect of the invention features an isolated, enriched, or purified nucleic acid molecule encoding one or more Yia operon-related polypeptides selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

In preferred embodiments, the isolated, enriched, or purified nucleic acid molecule encoding one or more Yia operon-related polypeptides comprises a nucleotide sequence that: (a) encodes a polypeptide having the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18; (b) is the complement of the nucleotide sequence of (a); and (c) hybridizes under highly stringent conditions to the nucleotide molecule of (a) and encodes a naturally occurring polypeptide.

In another preferred embodiment, the invention features an isolated, enriched, or purified nucleic acid molecule, wherein said nucleic acid molecule comprises the nucleotide sequence set forth in SEQ ID NO:19. The nucleic acid molecule comprises: (a) one or more nucleotide sequences that are set forth in SEQ ID NO:1, SEQ ID NO:2,

SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; (b) the complement of the nucleotide sequence of (a); (c) nucleic acid that hybridizes under stringent conditions to the nucleotide molecule of (a); (d) the full length sequence of SEQ ID NO:19, except that it lacks one or more of the sequences set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; or (e) is the complement of the nucleotide sequence of (d).

The term "complement" refers to two nucleotides that can form multiple thermodynamically favorable interactions with one another. For example, adenine is complementary to thymine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can form three hydrogen bonds. A nucleotide sequence is the complement of another nucleotide sequence if the nucleotides of the first sequence are complementary to the nucleotides of the second sequence. The percent of complementarity (*i.e.* how many nucleotides from one strand form multiple thermodynamically favorable interactions with the other strand compared with the total number of nucleotides present in the sequence) indicates the extent of complementarity of two sequences.

Various low or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired. These conditions are well-known to those skilled in the art. Under stringent hybridization

conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having 1 or 2 mismatches out of 20 contiguous nucleotides.

By "stringent hybridization conditions" is meant hybridization conditions at least as stringent as the following: hybridization in 50% formamide, 5X SSC, 50 mM NaH_2PO_4, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X Denhart's solution at 42 °C overnight; washing with 2X SSC, 0.1% SDS at 45 °C; and washing with 0.2X SSC, 0.1% SDS at 45 °C.

In other preferred embodiments the isolated, enriched, or purified nucleic acid molecule encoding one or more Yia operon-related polypeptides further comprises a vector or promoter effective to initiate transcription in a host cell. Preferably, the vector or promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI*^{q} repressor gene. In still other preferred embodiments, the nucleic acid molecule is isolated, enriched, or purified from a bacteria, preferably *Klebsiella oxytoca*.

The invention also features recombinant nucleic acid, preferably in a cell or an organism. The recombinant nucleic acid may contain a sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, or a functional derivative thereof, and a vector or a promoter effective to initiate transcription in a host cell. The

recombinant nucleic acid can alternatively contain a transcriptional initiation region functional in a cell, a sequence complementary to an RNA sequence encoding one or more Yia operon-related polypeptides and a transcriptional termination region functional in a cell.

In preferred embodiments, the isolated, enriched, purified, recombinant, or recombinant in a cell, nucleic acid comprises, consists essentially of, or consists of the full-length nucleic acid sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, encodes the full-length amino acid sequence of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18, a functional derivative thereof, or at least 35, 40, 45, 50, 60, 75, 100, 200, or 300 contiguous amino acids of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. The Yia operon-related polypeptides comprise, consist essentially of, or consist of at least 35, 40, 45, 50, 60, 75, 100, 200, or 300 contiguous amino acids of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. The nucleic acid may be isolated from a natural source by cDNA cloning or by subtractive hybridization. The natural source may be prokaryotic, eukaryotic, or protozoal, preferably bacterial, from the environment, and the nucleic

acid may be synthesized by the triester method or by using an automated DNA synthesizer. In other preferred embodiments, the nucleic acid molecule is isolated, enriched, or purified from a bacteria, preferably *Klebsiella oxytoca*.

In yet other preferred embodiments, the nucleic acid is a conserved or unique region, for example those useful for: the design of hybridization probes to facilitate identification and cloning of additional polypeptides, the design of PCR probes to facilitate cloning of additional polypeptides, obtaining antibodies to polypeptide regions, and designing antisense oligonucleotides.

By "conserved nucleic acid regions", are meant regions present on two or more nucleic acids encoding a Yia operon-related polypeptide, to which a particular nucleic acid sequence can hybridize under lower stringency conditions. Examples of lower stringency conditions are provided in Abe, *et al.* (J. Biol. Chem. 19:13361-13368, 1992), hereby incorporated by reference herein in its entirety, including any drawings, figures, or tables. Preferably, conserved regions differ by no more than 5 out of 20 nucleotides.

By "unique nucleic acid region" is meant a sequence present in a nucleic acid coding for a Yia operon-related polypeptide that is not present in a sequence coding for any other naturally occurring polypeptide. Such regions preferably encode 12 (preferably 15, more preferably 20,

most preferably 30) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; 30 (preferably 35, more preferably 40, most preferably 50) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; 5 (preferably 10, more preferably 15, most preferably 25) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; 17 (preferably 20, more preferably 25, most preferably 35) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; 11 (preferably 15, more preferably 20, most preferably 30) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:16. In particular, a unique nucleic acid region is preferably of bacterial origin.

A third aspect of the invention features a nucleic acid probe for the detection of nucleic acid encoding one or more Yia operon-related polypeptides, selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS, in a sample. Preferably, the nucleic acid probe encodes a polypeptide that is a fragment of the protein encoded by the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. The nucleic acid probe contains a nucleotide base sequence that will hybridize to the full-

length sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, or a functional derivative thereof. Hybridization is preferably under stringent conditions.

In preferred embodiments, the nucleic acid probe hybridizes to nucleic acid encoding at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; at least 30, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; at least 5, 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; at least 17, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; at least 11, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:16, or a functional derivative thereof.

Methods for using the probes include detecting the presence or amount of Yia operon-related RNA in a sample by contacting the sample with a nucleic acid probe under conditions such that hybridization occurs and detecting the presence or amount of the probe bound to Yia operon-related RNA. The nucleic acid duplex formed between the probe and a

nucleic acid sequence coding for a Yia operon-related polypeptide may be used in the identification of the sequence of the nucleic acid detected (Nelson *et al.*, in Non-isotopic DNA Probe Techniques, Academic Press, San Diego, Kricka, ed., p. 275, 1992, hereby incorporated by reference herein in its entirety, including any drawings, figures, or tables). Kits for performing such methods may be constructed to include a container means having disposed therein a nucleic acid probe.

A fourth aspect of the invention features a recombinant cell comprising a nucleic acid molecule encoding one or more Yia operon-related polypeptides selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS. In such cells, the nucleic acid may be under the control of the genomic regulatory elements, or, preferably, may be under the control of exogenous regulatory elements including an exogenous promoter. By "exogenous" is meant a promoter that is not normally coupled *in vivo* transcriptionally to the coding sequence for the Yia operon-related polypeptides.

In preferred embodiments, the recombinant cell comprises nucleic acid encoding a polypeptide that is a fragment of the protein encoded by the amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. By "fragment," is meant an amino acid sequence present in a Yia operon polypeptide.

Preferably, such a sequence comprises at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; at least 30, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; at least 5, 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; at least 17, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; at least 11, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:16.

Alternatively, the recombinant cell comprises the nucleic acid sequence set forth in SEQ ID NO:19, or comprises: (a) one or more nucleotide sequences that are set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; (b) the complement of the nucleotide sequence of (a); (c) nucleic acid that hybridizes under stringent conditions to the nucleotide molecule of (a); (d) the full length sequence of SEQ ID NO:19, except that it lacks one or more of the sequences set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; and (e) is the complement

of the nucleotide sequence of (d). Preferably, the recombinant cell further comprises a vector or promoter effective to initiate transcription of the above-identified nucleic acid in the cell. Preferably, the vector or promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI*^{q} repressor gene. Preferably, the recombinant cell is a bacteria, more preferably *Klebsiella oxytoca*.

Other preferred embodiments of this aspect of the invention include a recombinant cell useful for screening for one or more nucleic acid sequences that express one or more products that convert a source compound into a target compound, where the cell expresses one or more genes, comprising an inducible promoter, and where the one or more genes encodes one or more proteins that in the presence of the target compound and an inducer provide a detectable signal, where the detectable signal indicates the presence of the one or more nucleic acid sequences. Preferably, the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color, and most preferably is growth.

In preferred embodiments, of the recombinant cell useful for screening, the one or more nucleic acid sequences encodes a metabolic pathway not normally present in said cell. In other preferred embodiments, the nucleic acid is selected from the group consisting of mutagenized DNA, environmental DNA, combinatorial libraries, and recombinant

DNA. Preferably, the environmental DNA is selected from the group consisting of mud, soil, sewage, flood control channels, sand, and water. Preferably the mutagenized DNA is the result of enzyme mutagenesis where the mutagenesis is selected from the group consisting of random, chemical, PCR-based, and directed mutagenesis. The directed mutagenesis is to include, for example, DNA shuffling. Preferably the enzymes to be mutagenized in this way are selected from the group consisting of lactonases, esterhydrolases, and reductases.

Additionally in this preferred embodiment, the cell preferably requires the presence of the target compound and the inducer for growth. Preferably, the target compound is selected from the group consisting of ascorbate and 2-KLG. In addition, the one or more genes are preferably under the control of an inducible promoter, preferably comprising the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI*^{q} repressor gene. Preferably, the one or more proteins encoded by the one or more genes are one or more Yia operon-related polypeptides. Preferably, the cell naturally expresses the one or more genes, or has been genetically manipulated to express the one or more genes. Preferably, the cell is a bacteria, most preferably *Klebsiella oxytoca*.

A fifth aspect of the invention features one or more isolated, enriched, or purified Yia operon-related

polypeptides selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

By "isolated" in reference to a polypeptide is meant a polymer of 6 (preferably 12, more preferably 18, most preferably 25, 32, 40, or 50) or more amino acids conjugated to each other, including polypeptides that are isolated from a natural source or that are synthesized. In certain aspects longer polypeptides are preferred, such as those with 100, 200, 300, 400, or more contiguous amino acids of the sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17 or SEQ ID NO:18.

The isolated polypeptides of the present invention are unique in the sense that they are not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only amino acid chain present, but that it is essentially free (about 90-95% pure at least) of no-amino acid-based material naturally associated with it.

By the use of the term "enriched" in reference to a polypeptide is meant that the specific amino acid sequence constitutes a significantly higher fraction (2-5 fold) of the total amino acid sequences present in the cells or solution of interest than in normal or diseased cells or in

the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other amino acid sequences present, or by a preferential increase in the amount of the specific amino acid sequence of interest, or by a combination of the two. However, it should be noted that enriched does not imply that there are no other amino acid sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term significant here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other amino acid sequences of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no amino acid sequence from other sources. The other source of amino acid sequences may, for example, comprise amino acid sequence encoded by a yeast or bacterial genome, or a cloning vector such as pUC19. The term is meant to cover only those situations in which man has intervened to increase the proportion of the desired amino acid sequence.

It is also advantageous for some purposes that an amino acid sequence be in purified form. The term "purified" in reference to a polypeptide does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment. Compared to the natural level this level should be at least 2-5 fold

greater (*e.g.*, in terms of mg/mL). Purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. The substance is preferably free of substances present in its natural environment at a functionally significant level, for example 90%, 95%, or 99% pure.

In preferred embodiments, the polypeptide is a fragment of the protein encoded by the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. Preferably, the Yia operon polypeptide contains at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; at least 30, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; at least 5, 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; at least 17, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; at least 11, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:16, or a functional derivative thereof.

The polypeptide can be isolated from a natural source by methods well-known in the art. The natural source may be protozoal, eukaryotic, or prokaryotic, and the polypeptide may be synthesized using an automated polypeptide synthesizer. Preferably, the polypeptide is isolated, enriched, or purified from bacteria, most preferably *Klebsiella oxytoca*.

In some embodiments the invention includes one or more recombinant Yia operon-related polypeptides. By "recombinant Yia operon-related polypeptide" is meant a polypeptide produced by recombinant DNA techniques such that it is distinct from a naturally occurring polypeptide either in its location (*e.g.*, present in a different cell or tissue than found in nature), purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

In a sixth aspect, the invention features an antibody (*e.g.*, a monoclonal or polyclonal antibody) having specific binding affinity to a Yia operon-related polypeptide or a Yia operon-related polypeptide fragment. In preferred embodiments, the yia operon-related polypeptide is selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

By "specific binding affinity" is meant that the antibody binds to the target Yia operon-related polypeptide with greater affinity than it binds to other polypeptides under specified conditions. Antibodies or antibody

fragments are polypeptides which contain regions that can bind other polypeptides. The term "specific binding affinity" describes an antibody that binds to a Yia operon polypeptide with greater affinity than it binds to other polypeptides under specified conditions.

The term "polyclonal" refers to antibodies that are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen or an antigenic functional derivative thereof. For the production of polyclonal antibodies, various host animals may be immunized by injection with the antigen. Various adjuvants may be used to increase the immunological response, depending on the host species.

"Monoclonal antibodies" are substantially homogenous populations of antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. Monoclonal antibodies may be obtained by methods known to those skilled in the art (Kohler et al., Nature 256:495-497, 1975, and U.S. Patent No. 4,376,110, both of which are hereby incorporated by reference herein in their entirety including any figures, tables, or drawings).

The term "antibody fragment" refers to a portion of an antibody, often the hypervariable region and portions of the surrounding heavy and light chains, that displays specific binding affinity for a particular molecule. A

hypervariable region is a portion of an antibody that physically binds to the polypeptide target.

Antibodies or antibody fragments having specific binding affinity to a Yia operon-related polypeptide of the invention may be used in methods for detecting the presence and/or amount of Yia operon polypeptide in a sample by probing the sample with the antibody under conditions suitable for Yia operon-related-antibody immunocomplex formation and detecting the presence and/or amount of the antibody conjugated to the Yia operon-related polypeptide. Diagnostic kits for performing such methods may be constructed to include antibodies or antibody fragments specific for the Yia operon-related polypeptide as well as a conjugate of a binding partner of the antibodies or the antibodies themselves.

An antibody or antibody fragment with specific binding affinity to a Yia operon-related polypeptide of the invention can be isolated, enriched, or purified from a prokaryotic or eukaryotic organism. Routine methods known to those skilled in the art enable production of antibodies or antibody fragments, in both prokaryotic and eukaryotic organisms. Purification, enrichment, and isolation of antibodies, which are polypeptide molecules, are described above.

Antibodies having specific binding affinity to a Yia operon-related polypeptide of the invention may be used in methods for detecting the presence and/or amount of Yia

operon-related polypeptide in a sample by contacting the sample with the antibody under conditions such that an immunocomplex forms and detecting the presence and/or amount of the antibody conjugated to the Yia operon-related polypeptide. Diagnostic kits for performing such methods may be constructed to include a first container containing the antibody and a second container having a conjugate of a binding partner of the antibody and a label, such as, for example, a radioisotope. The diagnostic kit may also include notification of an FDA approved use and instructions therefor.

In a seventh aspect, the invention features a hybridoma that produces an antibody having specific binding affinity to a Yia operon-related polypeptide or a Yia operon-related polypeptide fragment. By "hybridoma" is meant an immortalized cell line that is capable of secreting an antibody, for example an antibody to a Yia operon-related polypeptide of the invention. In preferred embodiments, the antibody to the Yia operon-related polypeptide comprises a sequence of amino acids that is able to specifically bind a Yia operon-related polypeptide of the invention.

In an eighth aspect, the invention features a Yia operon-related polypeptide binding agent able to bind to a Yia operon-related polypeptide. The binding agent is preferably a purified antibody that recognizes an epitope present on a Yia operon-related polypeptide of the invention. Other binding agents include molecules that bind

to Yia operon-related polypeptides and analogous molecules which bind to a Yia operon-related polypeptide. Such binding agents may be identified by using assays that measure Yia operon-related binding partner activity, such as those that measure growth or ascorbate metabolism.

The invention also features a method for screening for other organisms containing a Yia operon-related polypeptide of the invention or an equivalent sequence. The method involves identifying the novel polypeptide in other organisms using techniques that are routine and standard in the art, such as those described herein for identifying the Yia operon-related polypeptide of the invention or others standard in the art (*e.g.*, cloning, Southern or Northern blot analysis, *in situ* hybridization, PCR amplification, etc.).

A ninth aspect of the invention features a method for identifying a substance that converts a source compound to a target compound, comprising: contacting a cell with nucleic acid, where the nucleic acid expresses a product that converts a source compound into a target compound, and where the cell expresses one or more proteins which in the presence of the target compound provide a detectable signal; contacting the cell with a test substance; and monitoring the detectable signal, where the detectable signal indicates the presence of the substance.

In preferred embodiments of the method for identifying a substance that converts a source compound to a

target compound, the substance is selected from the group consisting of antibodies, small organic molecules, peptidomimetics, and natural products. In other preferred embodiments, the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color. Preferably, the detectable signal is growth, and the target compound is metabolizable to an element selected from the group consisting of carbon, nitrogen, sulfur, and phosphorous, most preferably carbon. Alternatively, the target compound is metabolizable to an essential nutrient. In still other preferred embodiments of the invention, the source compound is selected from the group consisting of 2-KLG, 2,5-DKG, L-IA, L-GuA, and glucose.

In other highly preferred embodiments of the method for identifying a substance that converts a source compound to a target compound, the one or more proteins are one or more Yia operon-related polypeptides. Preferably, the Yia operon further comprises a vector or promoter effective to initiate transcription in a host cell, and most preferably the vector or promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI*^{q} repressor gene.

A tenth aspect of the invention features a method for detecting the presence, absence, or amount of a compound in a sample comprising: contacting the sample with a cell, where the cell expresses one or more genes encoding one or more proteins that in the presence of the compound provide a

detectable signal that indicates the presence, absence, or amount of said compound. A schematic of an example of a preferred embodiment of the method is shown in Fig. 13. In preferred embodiments, the compound is ascorbate and the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color. In other preferred embodiments, the one or more genes comprises *yiaJ*, and preferably further comprises a promoter transcriptionally linked to a reporter gene. Preferably, YiaJ is naturally expressed in the cell, or the cell has been genetically manipulated to express YiaJ. Preferably the reporter gene has a promoter transcriptionally linked and the expression of the reporter gene is regulated by the binding of YiaJ to the promoter. The binding of YiaJ to the promoter is preferably regulated by the presence or absence of ascorbate. Preferably the cell is a bacteria, and most preferably *Klebsiella oxytoca*.

An eleventh aspect of the invention features an isolated, purified, or enriched nucleic acid molecule encoding YiaJ and a reporter gene. Preferably, the nucleic acid molecule further comprises a promoter transcriptionally linked to a reporter gene. Preferably the reporter gene is regulated by the binding of YiaJ to the promoter. The binding of YiaJ to the promoter is preferably regulated by the presence or absence of ascorbate. In preferred embodiments, the nucleic acid molecule further comprises a

vector or promoter effective to initiate transcription in a host cell.

A twelfth aspect of the invention features a recombinant cell comprising the nucleic acid molecule described in the eleventh aspect of the invention, above.

Preferred embodiments of this aspect of the invention feature a recombinant cell for detecting the presence, absence, or amount of a compound in a sample, where the cell expresses one or more genes encoding one or more proteins that in the presence of the compound provide a detectable signal, where the signal indicates the presence, absence, or amount of the compound. In preferred embodiments, the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color.

In other preferred embodiments of the recombinant cell for detecting the presence, absence, or amount of a compound in a sample, the one or more genes comprises *yiaJ*, and further comprises a promoter transcriptionally linked to a reporter gene. Preferably, the expression of the reporter gene is regulated by the binding of YiaJ to the promoter. Preferably, *yiaJ* is naturally expressed in the recombinant cell, or the cell has been genetically manipulated to express *yiaJ*. The recombinant cell is preferably a bacteria, and more preferably *Klebsiella oxytoca*.

A thirteenth aspect of the invention features a method of selection for one or more nucleic acid sequences encoding a metabolic pathway from a source compound to a

target compound comprising: (1) identifying an organism that metabolizes a target compound to provide an essential element; (2) identifying one or more genes responsible for the metabolism of the target compound to the essential element; (3) expressing the one or more genes under the control of an inducible promoter, whereby the target compound is metabolized only in the presence of an inducer and not in the absence of the inducer; (4) expressing nucleic acid sequences potentially encoding the metabolic pathway in the recipient organism; and (5) selecting the recipient organism for growth in the presence of the source compound in the absence of the target compound and in the presence of the inducer, where growth on the source compound in the absence of the target compound and in the presence of the inducer indicates the presence of the nucleic acid sequence.

In preferred embodiments of the method of selection, the essential element is selected from the group consisting of carbon, phosphorous, nitrogen, and sulfur, and most preferably is carbon.

In other preferred embodiments, the method of selection further comprises the transfer of the one or more genes to a highly genetically manipulatable recipient organism, such that the recipient organism metabolizes the target compound to provide an essential element.

By a "highly genetically manipulatable recipient organism" is meant an organism, preferably single-celled,

more preferably bacteria, and most preferably *Klebsiella oxytoca*, that can be manipulated by the standard genetic techniques, including but not limited to, transfection, selection in selective media, growth in culture.

The summary of the invention described above is not limiting and other features and advantages of the invention will be apparent from the following detailed description of the invention, and from the claims.

DESCRIPTION OF THE FIGURES

Figure 1 shows a physical map of the *yiaK-S* operon, which includes the open reading frames *yiaK*, *yiaL*, *orf1*, *yiaX2*, *lyxK*, *yiaQ*, *yiaR*, and *yia*, and its putative regulator, *yiaJ*, compared with the *E. coli yiaK-S* operon, which includes the open reading frames *yiaK*, *yiaL*, *yiaM*, *yiaN*, *yiaO*, *lyxK*, *yiaQ*, *yiaR*, and *yiaS*, and its putative regulator *yiaJ*.

Figures 2A, 2B, 2C, 2D, 2E, and 2F show the nucleic acid sequence (SEQ ID NO:19) and translated amino acid sequences of the open reading frames of the yia operon and its putative regulator, *yiaJ*.

Figure 3 shows a multiple sequence alignment of YiaJ-Ko (SEQ ID NO:10), YiaJ-Ec (SEQ ID NO:20), and YiaJ-Hi (SEQ ID NO:21). Identical sequences among the three proteins are indicated by shading.

Figure 4 shows a multiple sequence alignment of YiaK-Ko (SEQ ID NO:11), YiaK-Ec (SEQ ID NO:22), and YiaK-Hi

(SEQ ID NO:23). Identical sequences among the three proteins are indicated by shading.

Figure 5 shows a multiple sequence alignment of YiaJ-Ko (SEQ ID NO:12), YiaL-Ec (SEQ ID NO:24), and YhcH-Hi (SEQ ID NO:25). Identical sequences among the three proteins are indicated by shading.

Figure 6 shows a multiple sequence alignment of LyxK-Ko (SEQ ID NO:15), LyxK-Ec (SEQ ID NO:26), and LyxK-Hi (SEQ ID NO:27). Identical sequences among the three proteins are indicated by shading.

Figure 7 shows a multiple sequence alignment of YiaQ-Ko (SEQ ID NO:16), YiaQ-Ec (SEQ ID NO: 28), and YiaQ-Hi (SEQ ID NO:29). Identical sequences among the three proteins are indicated by shading.

Figure 8 shows a multiple sequence alignment of YiaR-Ko (SEQ ID NO:17), YiaR-Ec (SEQ ID NO:30), and YiaR-Hi (SEQ ID NO:31). Identical sequences among the three proteins are indicated by shading.

Figure 9 shows a multiple sequence alignment of YiaS-Ko (SEQ ID NO:18), YiaS-Ec (SEQ ID NO:32), and YiaS-Hi (SEQ ID NO:33). Identical sequences among the three proteins are indicated by shading.

Figure 10 shows a schematic of the construction of the Tester Strain. The plasmid pMG125 is shown which comprises: (i) a chloramphenicol resistance marker (*cat*); (ii) the thermosensitive origin of replication from plasmid pHO1 (*pHO1 rep* (t^s)); (iii) a 0.8 kb fragment containing the

5′ region of the *yiaJ* gene and its promoter sequences; (iv) the spectinomycin resistance marker (*spc*); (v) the *lacI*^{q}-*lacO-trc* promoter fragment; and (vi) a 1 kb fragment containing the 5′ end of *yiaK*, including its ribosome binding site for translation initiation while excluding the promoter sequences of the *yiaK-S* operon. The recombinant plasmid pMG125 was introduced into *K. oxytoca* wild type strain VJSK009 by transformation at 30 °C, the permissive temperature for pMAK705 replication. Chromosomal integration of the pMG125 insert into VJSK009 was achieved by double crossover at the *yiaJ-K* locus such that the endogenous promoter of the *yiaK-S* operon was replaced with the inducible *lacI*^{q}-*trc* promoter system in the resulting recombinant cell, MGK003.

Figure 11 shows a schematic representation of a general example of a metabolic selection process. Briefly, genetic material, isolated from microbes, is incorporated into a Tester Strain and the gene(s) of interest selected for by growth on "S". The gene(s) of interest will catalyze the conversion of "S" to "T" in the Tester Strain, thereby allowing growth on "S".

Figure 12 shows a schematic representation of a more specific example of metabolic selection process, in which "S" is 2-KLG and "T" is AsA. In this case, the gene(s) of interest are those that catalyze the conversion of 2-KLG to AsA.

Figure 13, part A shows a theoretical model for AsA-dependent activation of the *yiaK-S* operon. Based on transcriptional analyses, the YiaJ regulatory protein is thought to activate transcription of the *yiaK-S* AsA catabolic operon in response to AsA present in the medium. However, the inventors do not wish to be held to this interpretation of the data.

Figure 13, part B shows a schematic representation of a whole-cell reporter system for AsA sensing. The *yiaK-S* promoter region (P_{yia}) is fused to the Green-Fluorescent-Protein (GFP) gene (or to *lux* or other reporter genes), and the fusion is integrated into the chromosome of an indicator strain, which also contains the YiaJ regulator. In the presence of AsA, YiaJ is stimulated and activates transcription of the *yia-GFP* fusion, thereby conferring an easily detectable GFP-positive or fluorescent phenotype.

DETAILED DESCRIPTION OF THE INVENTION

The instant invention is based in part on the use of a metabolic selection strategy that uses a recombinant DNA selection procedure to identify enzymatic pathways for the conversion of a source compound to a target compound. This technique allows at least a million-fold increase in the discovery rate over classical biochemical screening approaches, and allows testing of the 99% of the environmental microbes that are currently not able to be cultured in the laboratory.

The general process involves the creation/ identification of an easily genetically-manipulatable organism containing an inducible signal, such that the signal is activated when a target compound is metabolized, followed by the screening of nucleic acid in this organism to identify genes which metabolize a source compound to the target compound (Figs. 11 and 12)

In a specific embodiment, the process involves three steps (1) the identification of an organism capable of metabolizing the target compound to carbon and energy, and the transfer of this metabolic pathway to a highly genetically manipulatable organism, *e.g. Escherichia coli* or *Bacillus subtilis,* with the result that the recipient now uses the target compound for growth; (2) placing the expression of the pathway under the control of an inducible promoter, whereby the target compound is metabolized in the presence of an inducer and not in its absence; and (3) cloning genes, which are to be tested for their ability to metabolize the source compound, into the recipient, and selecting for growth on the source compound in the presence of the inducer but in the absence of the target compound.

Once positive organisms are identified in the above selection scheme by growth in the presence of inducer, the organisms are further screened for their ability to grow in the absence of the inducer. No growth in the absence of the inducer indicates that the metabolism of the source compound proceeds via the target compound. Thus, the

nucleic acid probably encodes an enzymatic pathway for the conversion of the source compound to the target compound.

Growth in the absence of the inducer indicates that metabolism of the source compound to the essential element or factor does not require prior conversion to the target compound, rather it may proceed directly, or through an intermediate, to the essential element or factor. When conversion directly to the target compound is the desired result, further work is necessary to obtain the desired genes. methods of obtaining the desired genes include: re-selection of DNA from other sources; random mutation of the DNA followed by re-selection; knocking out (deleting or blocking the expression of genes by methods well-known in the art) the genes that allow the direct conversion to the essential element or factor or from an intermediate to the essential element or factor followed by re-selection; etc. In one preferred embodiment, expression of the genes that allow the direct, or partially direct, conversion to the essential factor are knocked out or their expression blocked, thereby "forcing" the conversion to the essential element through the target compound. This will be effective if a pathway through the target compound existed, but was thermodynamically unfavorable, for example.

Alternatively, if the intermediate is freely interconvertable with the desired target compound as well as to the essential element, growth in the absence of the inducer may be an acceptable outcome, or even desirable. By

"freely interconvertable" is meant that an enzymatic pathway is present to allow the intermediate to be converted to the target. The interconvertability of the compounds would also be determined using the methods described above for obtaining a pathway directly to the target compound.

Under some circumstances, selection of a pathway directly, or through an intermediate, to the essential element or factor rather than to the target compound, is a preferred result. For example, under circumstances where the desired target compound is not one that can be used for direct selection (*e.g.* does not cross membranes or is rapidly broken down) a "surrogate target" might have to be used. A surrogate target refers to one that is used for selection, but is not the most highly desired target. In this embodiment, the target would preferably be on the pathway of conversion of the surrogate target to the essential element.

I. Functional Derivatives

Provided herein are functional derivatives of a polypeptide or nucleic acid of the invention. By "functional derivative" is meant a "chemical derivative," "fragment," or "variant," of the polypeptide or nucleic acid of the invention, which terms are defined below. A functional derivative retains at least a portion of the function of the protein, for example reactivity with an antibody specific for the protein, enzymatic activity or

binding activity mediated through noncatalytic domains, which permits its utility in accordance with the present invention. It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide which retains the functionality of the original. In both cases, all permutations are intended to be covered by this disclosure.

Also included with "functional derivatives" of the polypeptides, in particular, of the invention are "chemical derivatives". A "chemical derivative" contains additional chemical moieties not normally a part of the protein. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues, for example, as described below.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl

disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing primary amine containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine alpha-amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimide (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residue are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutaminyl and asparaginyl residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Derivatization with bifunctional agents is useful, for example, for cross-linking the component peptides of the protein to each other or to other proteins in a complex to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, for example, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-

dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl) dithiolpropioimidate yield photo-activatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (Creighton, T.E., Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half-life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein complex and the like. Moieties capable of mediating such effects are disclosed, for example, in Remington's Pharmaceutical Sciences, 18th ed., Mack Publishing Co., Easton, PA (1990).

The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the

proteins, of the complexes having a length less than the full-length polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably, the fragment is obtained recombinantly by appropriately modifying the DNA sequence encoding the proteins to delete one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence. Fragments of a protein are useful for screening for compounds that act to modulate enzyme activity, as described herein. It is understood that such fragments may retain one or more characterizing portions of the native complex. Examples of such retained characteristics include: catalytic activity; substrate specificity; interaction with other molecules in the intact cell; regulatory functions; or binding with an antibody specific for the native complex, or an epitope thereof.

Another functional derivative intended to be within the scope of the present invention is a "variant" polypeptide which either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The variant may be derived from a naturally occurring complex component by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify codons for one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence. It is understood that such variants

having added, substituted and/or additional amino acids retain one or more characterizing portions of the native protein, as described above.

A functional derivative of a protein with deleted, inserted and/or substituted amino acid residues may be prepared using standard techniques well-known to those of ordinary skill in the art. For example, the modified components of the functional derivatives may be produced using site-directed mutagenesis techniques (as exemplified by Adelman et al., 1983, DNA 2:183) wherein nucleotides in the DNA coding the sequence are modified, and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above. Alternatively, proteins with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art. The functional derivatives of the proteins typically exhibit the same qualitative biological activity as the native proteins.

II. Nucleic Acid Probes, Methods, and Kits for Detection of Yia operon-related polypeptides

A nucleic acid probe of the present invention may be used to probe an appropriate chromosomal or cDNA library by usual hybridization methods to obtain other nucleic acid molecules of the present invention. A chromosomal DNA or cDNA library may be prepared from appropriate cells

according to recognized methods in the art (cf. "Molecular Cloning: A Laboratory Manual", second edition, Cold Spring Harbor Laboratory, Sambrook, Fritsch, & Maniatis, eds., 1989).

In the alternative, chemical synthesis can be carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the amino acid sequence of the polypeptide of interest. The synthesized nucleic acid probes may be used as primers in a polymerase chain reaction (PCR) carried out in accordance with recognized PCR techniques, essentially according to PCR Protocols, "A Guide to Methods and Applications", Academic Press, Michael, *et al.*, eds., 1990, utilizing the appropriate chromosomal or cDNA library to obtain the fragment of the present invention.

One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer alignment and sequence analysis known in the art ("Molecular Cloning: A Laboratory Manual", 1989, supra). The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be visualized using known methods.

The nucleic acid probes of the present invention include RNA, as well as DNA probes, such probes being generated using techniques known in the art. The nucleic acid probe may be immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, and acrylic resins, such as polyacrylamide and latex beads. Techniques for coupling nucleic acid probes to such solid supports are well known in the art.

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample which is compatible with the method utilized.

One method of detecting the presence of nucleic acids of the invention in a sample comprises (a) contacting said sample with the above-described nucleic acid probe under conditions such that hybridization occurs, and (b) detecting the presence of said probe bound to said nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art

as described above. Samples to be tested include but should not be limited to RNA samples extracted from environmental samples.

A kit for detecting the presence of nucleic acids of the invention in a sample comprises at least one container means having disposed therein the above-described nucleic acid probe. The kit may further comprise other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents include, but are not limited to radiolabelled probes, enzymatic labeled probes (horseradish peroxidase, alkaline phosphatase), and affinity labeled probes (biotin, avidin, or steptavidin). Preferably, the kit further comprises instructions for use.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the probe or primers used in the assay, containers which contain wash reagents (such as

phosphate buffered saline, Tris-buffers, and the like), and containers which contain the reagents used to detect the hybridized probe, bound antibody, amplified product, or the like. One skilled in the art will readily recognize that the nucleic acid probes described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

III. DNA Constructs Comprising Yia Operon-Related Nucleic Acid Molecules and Cells Containing These Constructs.

The present invention also relates to a recombinant DNA molecule comprising, 5' to 3', a promoter effective to initiate transcription in a host cell and the above-described nucleic acid molecules. In addition, the present invention relates to a recombinant DNA molecule comprising a vector and an above-described nucleic acid molecule. The present invention also relates to a nucleic acid molecule comprising a transcriptional region functional in a cell, a sequence complementary to an RNA sequence encoding an amino acid sequence corresponding to the above-described polypeptide, and a transcriptional termination region functional in said cell. The above-described molecules may be isolated and/or purified DNA molecules.

The present invention also relates to a cell or organism that contains an above-described nucleic acid molecule and thereby is capable of expressing a polypeptide. The polypeptide may be purified from cells which have been

altered to express the polypeptide. A cell is said to be "altered to express a desired polypeptide" when the cell, through genetic manipulation, is made to produce a protein which it normally does not produce or which the cell normally produces at lower levels. One skilled in the art can readily adapt procedures for introducing and expressing either genomic, cDNA, or synthetic sequences into either eukaryotic or prokaryotic cells.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains nucleotide sequences which contain transcriptional and translational regulatory information and such sequences are "operably linked" to nucleotide sequences which encode the polypeptide. An operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression may vary from organism to organism, but shall in general include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

If desired, the non-coding region 3' to the sequence encoding a Yia operon polypeptide of the invention may be obtained by the above-described methods. This region may be retained for its transcriptional termination regulatory sequences, such as termination and polyadenylation. Thus, by retaining the 3'-region naturally contiguous to the DNA sequence encoding a polypeptide of the invention, the transcriptional termination signals may be provided. Where the transcriptional termination signals are not satisfactorily functional in the expression host cell, then a 3' region functional in the host cell may be substituted.

Two DNA sequences (such as a promoter region sequence and a sequence encoding a polypeptide of the invention) are said to be operably linked if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of a gene sequence encoding a polypeptide of the invention, or (3) interfere with the ability of the gene sequence of a polypeptide of the invention to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence. Thus, to express a gene encoding a polypeptide of the invention, transcriptional and

translational signals recognized by an appropriate host are necessary.

The present invention encompasses the expression of a gene encoding a polypeptide of the invention (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Prokaryotic hosts are, generally, very efficient and convenient for the production of recombinant proteins and are, therefore, one type of preferred expression system for polypeptides of the invention. Prokaryotes most frequently are represented by various strains of *E. coli*. However, other microbial strains may also be used, including other bacterial strains.

In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Examples of suitable plasmid vectors may include pBR322, pUC18, pUC19 and the like; suitable phage or bacteriophage vectors may include γgt10, γgt11 and the like; and suitable virus vectors may include pMAM-neo, pKRC and the like. Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as *E. coli*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*, *Klebsiella*, and the like. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

To express a polypeptide of the invention (or a functional derivative thereof) in a prokaryotic cell, it is necessary to operably link the sequence encoding the polypeptide of the invention to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (*i.e.*, inducible or derepressible). Examples of constitutive promoters include the *int* promoter of bacteriophage λ, the *bla* promoter of the β-lactamase gene sequence of pBR322, and the *cat* promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage λ (P_L and P_R), the *trp*, *recA*, *λacZ*, *λacI*, and *gal* promoters of *E. coli*, the α-amylase (Ulmanen *et al.*, J. Bacteriol. 162:176-182, 1985) and the ς-28-specific promoters of *B. subtilis* (Gilman *et al.*, Gene Sequence 32:11-20, 1984), the promoters of the bacteriophages of *Bacillus* (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, Inc., NY, 1982), and *Streptomyces* promoters (Ward *et al.*, Mol. Gen. Genet. 203:468-478, 1986). Prokaryotic promoters are reviewed by Glick (Ind. Microbiot. 1:277-282, 1987), Cenatiempo (Biochimie 68:505-516, 1986), and Gottesman (Ann. Rev. Genet. 18:415-442, 1984).

Proper expression in a prokaryotic cell also requires the presence of a ribosome-binding site upstream of the gene sequence-encoding sequence. Such ribosome-binding sites are disclosed, for example, by Gold *et al.* (Ann. Rev.

Microbiol. 35:365-404, 1981). The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on the type of host cell used to express the gene. As used herein, "cell", "cell line", and "cell culture" may be used interchangeably and all such designations include progeny. Thus, the terms "transformants" or "transformed cells" include the primary subject cell and cultures derived therefrom, without regard to the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. However, as long as mutant progeny have the same functionality as that of the originally transformed cell, they are considered to be the same cell or cell-line.

Host cells which may be used in the expression systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the polypeptide of interest. Transcriptional initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical (such as metabolite) regulation.

A nucleic acid molecule encoding a polypeptide of the invention and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell

either as a nonreplicating DNA or RNA molecule, which may either be a linear molecule or a closed covalent circular molecule. Alternatively, permanent expression may occur through the integration of the introduced DNA sequence into the host chromosome or as a circular plasmid.

A vector may be employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, *e.g.*, antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Additional elements may also be needed for optimal synthesis of mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama (Mol. Cell. Biol. 3:280-289, 1983).

The introduced nucleic acid molecule can be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or

viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Preferred prokaryotic vectors include plasmids such as those capable of replication in *E. coli* (such as, for example, pBR322, ColE1, pSC101, pACYC 184, πVX; "Molecular Cloning: A Laboratory Manual", 1989, supra). Bacillus plasmids include pC194, pC221, pT127, and the like (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, NY, pp. 307-329, 1982). Suitable *Streptomyces* plasmids include p1J101 (Kendall *et al.*, J. Bacteriol. 169:4177-4183, 1987), and streptomyces bacteriophages such as ϕC31 (Chater *et al.*, In: Sixth International Symposium on Actinomycetales Biology, Akademiai Kaido, Budapest, Hungary, pp. 45-54, 1986). *Pseudomonas* plasmids are reviewed by John *et al.* (Rev. Infect. Dis. 8:693-704, 1986), and Izaki (Jpn. J. Bacteriol. 33:729-742, 1978).

Once the vector or nucleic acid molecule containing the construct(s) has been prepared for expression, the DNA construct(s) may be introduced into an appropriate host cell by any of a variety of suitable means, *i.e.*, transformation, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium

phosphate-precipitation, direct microinjection, and the like. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene(s) results in the production of a polypeptide of the invention, or fragments thereof. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of bromodeoxyuracil to neuroblastoma cells or the like). A variety of incubation conditions can be used to form the peptide of the present invention. The most preferred conditions are those which mimic physiological conditions.

V. Antibodies, Hybridomas, Methods of Use and Kits for Detection of Yia Operon-Related polypeptides

The present invention relates to an antibody having binding affinity to a polypeptide of the invention. The polypeptide may have the amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, or a functional derivative thereof, or at least 6 contiguous amino acids thereof (preferably, at least 15, 20, 25, 30, 35, or 40 contiguous amino acids thereof).

The present invention also relates to an antibody having specific binding affinity to a polypeptide of the invention. Such an antibody may be isolated by comparing its binding affinity to a polypeptide of the invention with

its binding affinity to other polypeptides. Those which bind selectively to a polypeptide of the invention would be chosen for use in methods requiring a distinction between a polypeptide of the invention and other polypeptides. Such methods could include, but should not be limited to, the identification of other cells expressing the polypeptides of the invention.

The polypeptides of the present invention can be used in a variety of procedures and methods, such as for the generation of antibodies, for use in identifying pharmaceutical compositions, and for selection of other enzymmatic pathways.

The polypeptides of the present invention can be used to produce antibodies or hybridomas. One skilled in the art will recognize that if an antibody is desired, such a peptide could be generated as described herein and used as an immunogen. The antibodies of the present invention include monoclonal and polyclonal antibodies, as well fragments of these antibodies.

The present invention also relates to a hybridoma which produces the above-described monoclonal antibody, or binding fragment thereof. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing monoclonal antibodies and hybridomas are well known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory

Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1984; St. Groth *et al.*, J. Immunol. Methods 35:1-21, 1980). Any animal (mouse, rabbit, and the like) which is known to produce antibodies can be immunized with the selected polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of polypeptide used for immunization will vary based on the animal which is immunized, the antigenicity of the polypeptide and the site of injection.

The polypeptide may be modified or administered in an adjuvant in order to increase the peptide antigenicity. Methods of increasing the antigenicity of a polypeptide are well known in the art. Such procedures include coupling the antigen with a heterologous protein (such as globulin or β-galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or

radioimmunoassay (Lutz *et al.*, Exp. Cell Res. 175:109-124, 1988). Hybridomas secreting the desired antibodies are cloned and the class and subclass are determined using procedures known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", supra, 1984).

For polyclonal antibodies, antibody-containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The above-described antibodies may be detectably labeled. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horse radish peroxidase, alkaline phosphatase, and the like) fluorescent labels (such as FITC or rhodamine, and the like), paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art, for example, *see* Stemberger *et al.*, J. Histochem. Cytochem. 18:315, 1970; Bayer *et al.*, Meth. Enzym. 62:308-, 1979; Engval *et al.*, Immunol. 109:129-, 1972; Goding, J. Immunol. Meth. 13:215-, 1976. The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues which express a specific peptide.

The above-described antibodies may also be immobilized on a solid support. Examples of such solid

supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10, 1986; Jacoby *et al.*, Meth. Enzym. 34, Academic Press, N.Y., 1974). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as immuno-chromatography.

Furthermore, one skilled in the art can readily adapt currently available procedures, as well as the techniques, methods and kits disclosed herein with regard to antibodies, to generate peptides capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides (Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides", In Synthetic Peptides, A User's Guide, W.H. Freeman, NY, pp. 289-307, 1992; Kaspczak *et al.*, Biochemistry 28:9230-9238, 1989).

Anti-peptide peptides can be generated by replacing the basic amino acid residues found in the peptide sequences of the Yia operon polypeptides of the invention with acidic residues, while maintaining hydrophobic and uncharged polar groups. For example, lysine, arginine, and/or histidine residues are replaced with aspartic acid or glutamic acid and glutamic acid residues are replaced by lysine, arginine or histidine.

The present invention also encompasses a method of detecting a Yia operon-related polypeptide in a sample, comprising: (a) contacting the sample with an above-described antibody, under conditions such that immunocomplexes form, and (b) detecting the presence of said antibody bound to the polypeptide. In detail, the methods comprise incubating a test sample with one or more of the antibodies of the present invention and assaying whether the antibody binds to the test sample. Detection of a polypeptide of the invention in a sample may indicate the presence of the pathway of the invention in other cells.

Conditions for incubating an antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the antibody used in the assay. One skilled in the art will recognize that any one of the commonly available immunological assay formats (such as radioimmunoassays, enzyme-linked immunosorbent assays, diffusion-based Ouchterlony, or rocket immunofluorescent assays) can readily be adapted to employ the antibodies of the present invention. Examples of such assays can be found in Chard ("An Introduction to Radioimmunoassay and Related Techniques" Elsevier Science Publishers, Amsterdam, The Netherlands, 1986), Bullock *et al.* ("Techniques in Immunocytochemistry," Academic Press, Orlando, FL Vol. 1, 1982; Vol. 2, 1983; Vol. 3, 1985), Tijssen ("Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in

Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1985).

The immunological assay test samples of the present invention include cells, protein or membrane extracts of cells, or environmental samples. The test samples used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can readily be adapted in order to obtain a sample which is testable with the system utilized.

A kit contains all the necessary reagents to carry out the previously described methods of detection. The kit may comprise: (i) a first container means containing an above-described antibody, and (ii) second container means containing a conjugate comprising a binding partner of the antibody and a label. Preferably, the kit also contains instructions for use. In another preferred embodiment, the kit further comprises one or more other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound antibodies.

Examples of detection reagents include, but are not limited to, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the chromophoric, enzymatic, or antibody binding reagents which

are capable of reacting with the labeled antibody. The compartmentalized kit may be as described above for nucleic acid probe kits. One skilled in the art will readily recognize that the antibodies described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

Other methods associated with the invention are described in the examples disclosed herein.

EXAMPLES

The examples below are not limiting and are merely representative of various aspects and features of the present invention. The examples below demonstrate the construction and use of metabolic selection systems, and the isolation of desired enzymatic pathways.

EXAMPLE 1: Construction of a Tester Strain for the Selection of Pathways from 2-KLG to AsA

This example is exemplary of how to construct tester strains, and therefore can be applied to the identification and construction of tester strains for the selection of other metabolic pathways. The basic idea is to take environmental samples and test them for growth on a target compound (in the example, ascorbate). Then, positive colonies are screened for the inability to grow on the source compound (in the example, 2-KLG). The tester strain is the one that grows on the target, but not the source

compound. Once the genes encoding the metabolic pathway for the target compound to the essential factor (an element such as carbon, nitrogen, sulphur or phosphorous, or a nutrient, for example) are identified, they are then place under the control of an inducible promoter, and the tester strain is ready to be utilized to select for the metabolic pathway from the source to the target compound.

If it proves difficult to obtain a tester strain that grows on the target, but not the source, but strains exist that do not grow on the source, then the pathway that permits growth on the target can be isolated and transferred to another strain that doesn't grow on the source in order to obtain the desired tester strain.

Isolation of a Strain that Grows on AsA, but not 2-KLG

Samples from diverse natural environments were collected to use for the isolation of microbes that can utilize ascorbic acid (AsA) as the sole carbon source. No bacterial species has previously been reported to grow on AsA minimal medium.

Environmental samples were collected from freshwater lakes, lemon and orange orchards, residential backyard soils, human and animal solid wastes.

Over 100 microbial isolates, capable of forming visible colonies within 20 hours of incubation at 30 °C on M9 minimal medium containing 0.5% AsA, were selected from these samples. These 100 isolates were then screened for

their ability to grow on 2-Keto-L-Gulonate (2-KLG) minimal medium.

One of the isolates that could utilize AsA as its sole source of carbon and energy, but could not grow on 2-KLG, was identified as *Klebsiella oxytoca* (Table 1). Thus, *Klebsiella oxytoca* was retained as a candidate for genetic engineering of a host strain that can use AsA under controlled conditions for the selection of cloned microbial pathways from 2-KLG to AsA.

Other bacterial strains capable of metabolizing ascorbic acid to carbon and energy were also identified, as were some that also metabolized 2KLG to carbon and energy (Table 1).

TABLE 1

COMPOUND UTILIZATION OF ENVIRONMENTAL ISOLATES

	AsA	**2-KLG**
GRAM POSITIVES	72 HR	24 HR
Bacillus megaterium	+	+
Streptomyces species	++	++
Yellow Bug	++	+++
GRAM NEGATIVES	24 HR	72 HR
Klebsiella pneumoniae	+++	-
Klebsiella species	+++	-

Klebsiella oxytoca	+++	-
Unknown Malodorous Short Rod	++	-

Identification of Genes Responsible for AsA Catabolism

In order to identify the gene(s) responsible for AsA catabolism in *K. oxytoca*, mutagenesis by transposition insertion was performed in *K. oxytoca* strain VJSK009 (Cali, B. M., *et al.*, 1989. J. Bacteriol. **171**:2666-2672) using the pfd-Tn5 delivery vector as described by Metzger, M., *et al.*, 1992. Nucl. Acids Res. **20**:2265-2270. Among 5,000 clones screened, several mutants that were no longer capable of growing on AsA were identified, most of which were also affected in their ability to grow on conventional carbon sources such as glucose, maltose, pyruvate or succinate. Two of the mutants, however, were specifically affected in AsA utilization and were further characterized by cloning and sequencing the regions adjacent to the transposon insertion.

Characterization of the Genes/Proteins of the Operon

In both mutants, the Tn5 insertion was found to disrupt the same operon of 8 genes. This operon was found to be homologous to the *yiaK-S* operon of *E. coli* (Blattner, F. R., *et al.*, 1997. Science **277**:1453-1462) which is thought to be involved with carbohydrate utilization (Badia, J., *et al.*, 1998. J. Biol. Chem. **273**:8376-8381).

Similarly to *E. coli*, the *K. oxytoca yiaK-S* operon is preceded by a transcriptional regulator, *yiaJ*. A

physical map of the *yiaK-S* operon and its putative regulator is shown in Figure 1. The nucleic acid sequence and translated amino acid sequence of the open reading frames of the operon and its putative regulator are shown in Figure 2 A-F.

The functions of the *yia* operon gene products in *K. oxytoca* and *E. coli* are unknown, except for the *E. coli lyxK*-encoded enzyme which was shown to phosphorylate L-xylulose and play a key role in the utilization of L-lyxose by *E. coli* (Sanchez, J. C., *et al.*, 1994. J. Biol. Chem. **169**:29665-29669). However, the *yiaK-S* operon is thought to be silent in wild-type *E. coli*, L-xylulose activity could not be detected in wild type cells, and *E. coli* K12 is unable to metabolize L-lyxose (Sanchez, J. C., *et al.*, 1994. *supra*). A similar operon is also present in *Haemophilus influenzae*, but no function has been determined for any of the open reading frames (Fleischmann, R.D., *et al.*, 1995. Science **269**:496-512).

Alignments of the yia open reading frames common among the three species are shown (Figs. 3-9). Based on sequence similarities, *yiaQ* has been classified as a putative hexulose-6-phosphate synthase, *yiaR* as a putative hexulose-6-phosphate isomerase, and *yiaS* as a putative sugar isomerase (data not shown).

Place Operon under the control of an Inducible Promoter

To engineer *K. oxytoca* as a host strain for the selection of biocatalysts which produce AsA, the promoter of the *yiaK-S* operon was replaced with a DNA fragment that contained the *trp-lac* hybrid promoter of transcription, the *lacO* operator, and the *lacI*^{q} repressor gene (Brosius, J. 1992. Meth. Enzymol. **216**:469-483). This allows the *yiaK-S* operon, and therefore AsA catabolism, to be turned ON and OFF in a tightly controlled manner in the presence or absence of IPTG, a non-metabolizable inducer of the *lac* promoter. Practically, a 5-way ligation was set up among: (i) the pMAK705 integration vector which carries a chloramphenicol resistance marker and the thermosensitive origin of replication from plasmid pHO1 (Hamilton, C. M., *et al.*, 1989. J. Bacteriol. **171**:4617-4622); (ii) a 0.8 kb fragment containing the 5′ region of the *yiaJ* gene and its promoter sequences; (iii) the spectinomycin resistance marker retrieved from *Staphylococcus aureus* Tn*554* (Murphy, E. 1985. Mol. Gen. Genet. **200**:33-39) to follow integration events; (iv) the *lacI*^{q}*-lacO-trc* promoter fragment retrieved from pSE380 (InVitrogen, Carlsbad, CA); and (v) a 1 kb fragment containing the 5′ end of *yiaK*, including its ribosome binding site for translation initiation while excluding the promoter sequences of the *yiaK-S* operon (Figure 10).

The recombinant plasmid, pMG125, was introduced into *K. oxytoca* wild type strain VJSK009 by transformation at 30 °C, the permissive temperature for pMAK705

replication. Chromosomal integration of the pMG125 insert by double crossover at the *yiaJ-K* locus was achieved by successive temperature switches as described by (Hamilton, C. M., *et al.*, 1989. *supra*). PCR analyses were performed on 12 candidates to verify that the endogenous promoter of the *yiaK-S* operon had been replaced with the inducible *lacI*^{q}*-trc* promoter system (Figure 10).

The resulting strain, MGK003, proved able to grow on M9 minimal medium supplemented with AsA 0.25% and IPTG 10 to 100 μM, while no growth was observed on the same medium lacking IPTG.

EXAMPLE 2: Preparation of Environmental DNA Libraries

An example of a currently preferred method for the isolation of DNA from environmental samples is provided below. In the example, purification from soil and water samples are described, however samples can be from any environmental source and the methods adapted according to practices well-known in the art.

Direct Isolation of Total DNA from Soil and Water Samples

Total microbial DNA was isolated from various soil and water samples according to the following procedure which is derived and modified from Steffan, R.J., *et al.*, 1988. Appl. Environ. Microbiol. **54**:2908-2915; Whatling, C. A., and C. M. Thomas. 1993. Anal. Biochem. **210**:98-101; and Zhou, J., *et al.*, 1996. Appl. Environ. Microbiol. **62**:316-322.

1. Begin with 100 g wet soil or 50 g dry soil; 150 mL sodium phosphate buffer 0.1 M, pH 4.5; and 5 g PVPP (acid washed).
2. Blender - medium speed - 3 times for 1 min (cool down between each cycle).
   Add 0.5 mL SDS 20%, blend 5 more seconds.
3. Centrifuge 10 min at 1,000 g at 10 °C.
4. Keep supernatant.
   Repeat extraction twice with soil pellet.
5. Combine the 3 supernatants.
   Centrifuge 20 min at 10,000 g at 10 °C
6. Wash pellet with cold 0.1% sodium-0.1% sodium pyrophosphate.
   Homogenize with blender for 1 min or shake.
   Centrifuge 20 min at 10,000 g at 10 °C.
7. Wash pellet with 33 mM Tris-HCl, 1 mM EDTA, pH 8.0.
8. Resuspend in 2 mL 10 mM Tris, pH 7.6; 1 N NaCl.
9. Mix with equal volume 1.2% LMP agarose at 42 °C.
   Pour into 1 mL syringes.
   Polymerize for 20 min at 4 °C.
10. Incubate 3-4 hours at 37 °C in 20 vol. 1 N NaCl; 100 mM EDTA; 10 mM Tris, pH 7.5; 1% sarkosyl; 1 mg/mL lysozyme.
11. Add 1 mg/mL proteinase K.
    Incubate overnight at 45 °C.
12. Wash agarose plugs twice with TE.

Store in 100 mM EDTA; 10 mM Tris at 4 °C.

13. Load noodles on LMP agarose gel 0.7%.

Cut out chromosomal band.

Heat 15 min at 65 °C in TE buffer.

Add 2 U GelZyme (InVitrogen) per 200 μL 1% agarose. Incubate for 2 h at 40 °C.

EtOH precipitate for no more than 30 min at -20 °C.

Preparation of Total DNA from Post-Enrichment Cultures

Aliquots from 18 water or soil samples were used to inoculate 50 mL of M9 minimal medium supplemented with any one of the following carbon sources: 0.5% 2-KLG; 0.25% L-idonate (L-IA); 0.25% L-gulonate (L-GuA) and 0.25% ascorbate. Culture flasks were incubated for 2 to 3 days at 30 °C without agitation.

Total DNA was isolated from these cultures as follows:

1. 20 mL were centrifuged for 5 min at 6,000 rpm.

2. Pellets were washed with 5 mL Tris 10 mM, EDTA 1 mM pH 8.0 (TE), were centrifuged again, and were resuspended in 0.9 mL TE.

3. Lysozyme (5 mg/mL) and RNase 100 (μg/mL) were added, and cells were incubated for 10 min at 37 °C.

4. Sodium dodecylsulfate (SDS) was added to a final concentration of 1%, and the tubes were gently shaken until lysis was completed.

5. 200 mL of a 5 N $NaClO_4$ stock solution were added to the lysate.

6. The mixture was extracted once with one volume of phenol:chloroform (1:1) and once with one volume of chloroform.

7. Chromosomal DNA was precipitated by adding 2 mL of cold (-20 °C) ethanol and gently coiling the precipitate around a curved Pasteur pipette.

8. DNA was dried for 30 min at room temperature and was resuspended in 100 to 500 μL of Tris 10 mM, EDTA 1 mM, NaCl 50 mM pH 8.0 to obtain a DNA concentration of 0.5 to 1 μg/μL.

EXAMPLE 3: Selection for Nucleic Acid which Converts 2-KLG to AsA (Fig. 12)

This example is exemplary of how to select for nucleic acid sequences that encode metabolic pathways, and therefore can be applied to the identification and selection of sequences encoding other metabolic pathways. Basically, a nucleic acid library is made, according to methods well-known in the art, from nucleic acid sequences isolated from environmental samples (as described in Example 2, for example). This library is then transfected into the tester strain and the resulting pool of transfected cells selected for growth on the source compound (2-KLG in the example) in the absence of the target compound (ascorbate in the example) and the presence of the inducer.

Construction of an Enrichment DNA Library in a Cosmid Vector

The SuperCos1 cosmid vector (Stratagene, La Jolla, CA) is a λ-based cloning system suitable for the cloning of large DNA fragments. After treatment according to the manufacturer's instructions, the 8 kb-long vector appears as two arms flanked by *cos* sites which are recognized by the λ-packaging machinery. Since only DNA molecules from 40 to 48 kb are efficiently packaged in λ-heads, this allows the selective cloning of 32 to 40 kb inserts between the two arms.

Chromosomal DNA extracted from 20 post-enrichment cultures was mixed in equal amounts. Five to ten µg of the mixture were partially digested with *Sau3A* restriction enzyme to obtain DNA fragments sized between 5 and 50 kb, were dephosphorylated, and were ligated with SuperCos1 arms using conditions recommended by the supplier. One µg of the ligation mixture was used in an *in vitro* packaging reaction using the Gigapack III Gold packaging kit from Stratagene to create the cosmid library.

Clearly, this procedure can be used to make other chromosomal DNA libraries, for example from other enriched environmental samples, or from chromosomal DNA extracted directly from environmental samples.

Transfection and Selection of the Cosmid Library

Prior to transfection of *K. oxytoca* strain MGK003 with the packaging mixture, the tester strain was transformed with plasmid pCB382 expressing the *E. coli lamB* gene that functions as λ receptor, which appears to be absent or non-functional in most *Klebsiella* strains (De Vries, G. E., *et al.*, 1984. Proc. Natl. Acad. Sci. USA **81**:6080-6084). The resulting MGK003 [λ^s] strain was transfected with the packaged products as follows:

1. Five mL of liquid LB medium supplemented with 0.2% maltose and 10 mM $MgSO_4$ were inoculated from an overnight preculture of strain MGK003 [pCB382].

2. Cells were grown to an OD_{600} of 0.5, were centrifuged at 500 xg for 10 min, and were resuspended in the same volume of 10 mM $MgSO_4$.

3. The packaging products were mixed with 2 mL of cells in 15 mL culture tubes, and were incubated for 20 min at 39 °C without shaking.

4. After adding 2.5 mL of 2x YT (1% NaCl; 1% yeast extract; 1.6% tryptone), cells were incubated at 37 °C for 1 h under gentle agitation.

5. A 100 μL-aliquot was plated on LB-kanamycin medium to determine the number of clones present in the cosmid library.

6. The remainder was centrifuged at 3000 g for 5 min and was resuspended in 1 mL of M9 minimal medium supplemented with 10 μM IPTG (IPTG concentration can be

varied up to 100 μM), and aliquots (200 μL) were plated on M9 plates containing 0.5% 2-KLG and 50 μM IPTG. 7.

Plates were incubated at 37 °C for 36 h for selecting candidate pathways that would convert 2-KLG to AsA. (Alternatively, selection can be done at 30 °C.)

Among 500,000 clones to which a first selection round was applied, approximately 100 colonies of various sizes appeared on 2-KLG/IPTG plates. These were re-streaked on: (i) LB-kanamycin to verify the presence of the cosmid vector; (ii) 2-KLG/IPTG; and (iii) 2-KLG lacking IPTG to determine if growth of the positive clones on 2-KLG was dependent upon the expression of AsA catabolism.

Two clones were retained that grew on LB-kanamycin and 2-KLG/IPTG, but not on 2-KLG without IPTG within 20 h at 37 °C. To verify that the observed phenotype was conferred by the cloned DNA, cosmid DNA was extracted from these two clones and introduced, by electroporation, into strain MGK003. In both cases, the back-cross gave a phenotype identical to that of the original clone obtained in the selection process (Data not shown).

Selection of libraries can also be done on other carbon sources to isolate other pathways, for example on L-gulonate (0.25%) plus IPTG to isolate pathways from L-gulonate to AsA, or on L-idonate (0.25%) plus IPTG to isolate pathways from L-idonate to AsA.

EXAMPLE 4: Isolation of Other Pathways

The metabolic selection strategy described above can also be used for the isolation of other pathways of interest, for example from 2-KLG to L-idonate, or 2-KLG to L-gulonate, or alternatively, to identify new reductase enzymes capable of the conversion of 2,5-DKG to 2-KLG. This conversion is one of the slow steps in the production of ascorbate, so identification of an enzymatic method would be economically useful. Basically, the strategy described in the examples above can be used to isolate any pathway to metabolize a compound as a carbon, nitrogen, sulfur, or potentially, a phosphorous source.

EXAMPLE 5: Directed Evolution of Enzymes

This metabolic selection method is also capable of facilitating the directed evolution of enzymes. One can use this technique to screen known enzymes for mutations leading to higher efficiency, or to better specify optimal temperature or cofactor requirements, in the metabolic utilization of a compound. The mutations can be the result of natural evolution, the result of PCR or chemical mutagenesis, or created through techniques like DNA shuffling.

EXAMPLE 6: Glucose to Ascorbic Acid Directly

Another permutation on this strategy that can be envisioned is to find new pathways for already existing

processes, *e.g.* selection for a new pathway for the conversion of glucose to ascorbic acid using only a few enzymatic steps. This is feasible using, for example, a strain for which the sequence of the entire genome is known, such as *E. coli* or *B. subtilis*. The genes for the metabolism of glucose can be mutagenized such that the strain can no longer use glucose as a carbon/energy source, and then glucose-utilization pathways can be selected for as described in the previous examples.

EXAMPLE 7: Ascorbate Biosensor (Fig. 13)

As mentioned above, the *yiaJ* protein is thought to be a regulator for the Yia operon. The experiments of the invention indicate that the regulatory activity of YiaJ may be, in part, modulated by sensing ascorbate. Thus, it is currently believed that the "sensing" of ascorbate by YiaJ (perhaps through binding, although the authors do not wish to be restricted to this interpretation) leads to the activation of the Yia operon, and thus the use of ascorbate as a carbon/energy source. This potentially results in an extremely sensitive "biosensor" for ascorbate. Thus, for example, it is envisioned that *yiaJ* could be placed in a construct such that when YiaJ bound ascorbate a detectable signal resulted, *i.e.* instead of turning "ON" or "OFF" the Yia operon, YiaJ could turn "ON" or "OFF" a gene which produces a detectable signal, for example a gene for fluorescence (*e.g.* β-galactosidase), luminescence (*e.g.*

luciferase), or color (lac operon, or green flourescent protein). Methods of constructing these signal constructs are well-known in the art (*e.g.* Simpson, *et al.* 1998. TIBTECH 16: 332-338; Applegate, *et al.* 1998. Applied Environ. Microbiol. 64: 2730-2735; Selifonova and Eaton, 1996. Applied Environ. Microbiol. 62: 778-783).

These biosensor constructs can also be used in the methods of the invention for screening for a metabolic selection pathway instead of using selection on an essential factor or element. In this case, the tester strain would be one that does not have the source to target pathway as determined by the absence of target being detected by the biosensor in the presence or the absence of the source compound. Thus, the biosensor would need to "sense" and to "react to" the presence of the target compound by any one of the methods described above. Following transfection of the library of nucleic acid from environmental sources, the resulting cells would be screened for the presence of the target compound using the biosensor. In order to facilitate the numbers of colonies that would need to be screened, this could be automated read in luminescent or flourescent readers or sorted by FACS prior to further testing and identification of individual colonies. Although this requires more initial screening than selection using an essential element, this method offers an alternative approach when the appropriate tester strain or the metabolic pathway is not available for screening using an essential

factor. Thus, the biosensor method provides the flexibility to identify pathways for compounds that are not metabolizable to an essential element, factor, or nutrient, but can be any compound for which a "biosensor" can be identified. Biosensors can be identified and created as described above.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, treatments, molecules, specific compounds described herein are presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not

specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising", "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.

Other embodiments are within the following claims.